

De psychometrische kwaliteit van drie toetsen voor basiscommunicatievaardigheden

J. Beullens, C. Aubry, H. Jaspaert

Samenvatting

Inleiding: In het kader van een evaluatie van een zelfstudiepakket voor basiscommunicatievaardigheden, bedoeld voor zowel geneeskundestudenten als studenten uit andere opleidingen tot hulpverlener, werden toetsen samengesteld om deze vaardigheden, onafhankelijk van medische inhoud, te meten.

Methode: Voor elk van de drie basisdimensies, 'luisteren', 'vragen stellen' en 'informer', werd een station ontwikkeld met een gestandaardiseerde gesprekspartner. Daarnaast werden vier schriftelijke toetsen en vier videotoetsen ontwikkeld. De toetsen werden afgenomen bij geneeskundestudenten uit het vierde jaar, huisartsgeneeskundestudenten (uit het zevende jaar) en farmaciestudenten (uit het vierde jaar). Na itemanalyse resteerden drie versies van zowel de schriftelijke als de videotoetsen, zodat in totaal sprake was van zeven toetsen.

Resultaten: De stationstoets bleek betrouwbaarder te zijn dan de andere toetsen. De huisartsgeneeskundestudenten, die een communicatietraining hadden gevolgd, behaalden op vijf van de zeven toetsen een significant hogere score dan de geneeskundestudenten, die geen communicatietraining hadden gevolgd. De stationstoets, schriftelijke toetsen en videotoetsen correleerden onderling slechts beperkt.

Conclusies: De hogere scores van de studenten huisartsgeneeskunde vormen een aanwijzing voor de validiteit van de toetsen. De beperkte onderlinge correlatie duidt erop dat de toetsen verschillende aspecten van communicatievaardigheid meten. Een betrouwbare en valide beoordeling van basiscommunicatievaardigheden moet daarom uit verschillende toetsen bestaan. De toetsen meten algemene communicatievaardigheden. Om medische communicatievaardigheden te meten zouden vergelijkbare toetsen gebruikt kunnen worden waarvan de inhoud betrekking heeft op verschillende medische probleemsituaties. (Beullens J, Aubry C, Jaspaert H. De psychometrische kwaliteit van drie toetsen voor basiscommunicatievaardigheden. Tijdschrift voor Medisch Onderwijs 2001;20(2):54-60.)

Inleiding

In de Faculteit Geneeskunde van de Katholieke Universiteit Leuven werd een pakket voor begeleide zelfstudie van basiscommunicatievaardigheden ontwikkeld. Aangezien het de bedoeling is dit pakket vroeg in de medische opleiding aan te bieden en bruikbaar te maken voor andere opleidingen van hulpverleners, werd het onafhankelijk van de medische inhoud samengesteld. Om dit pakket te evalueren werden toetsen ontwikkeld voor de meting

van drie vaardigheden: luisteren, vragen stellen en informeren. In dit onderzoek wordt gerapporteerd over de psychometrische kwaliteit van deze toetsen.

De meest voor de hand liggende wijze om adequaat gespreksvaardigheden te meten, is door middel van een stationstoets met behulp van gestandaardiseerde gesprekspartners. Dit zijn personen die getraind zijn om herhaaldelijk accuraat en consistent een bepaalde rol te spelen.¹ In het medisch onderwijs wordt de stationsproef met

simulatiepatiënten dikwijls gebruikt om het effect van een communicatietraining te meten. In onderzoek betreft het doorgaans slechts één station, dat vóór en na een training of bij een experimentele en een controlegroep wordt ingezet.²⁻⁷ Een stationstoets voor communicatievaardigheden van geneeskundestudenten met meerdere stations is zeldzaam.⁸ Een korte stationstoets met gestandaardiseerde gesprekspartners heeft in de regel een lage betrouwbaarheid tengevolge van de geringe generaliseerbaarheid van de bevindingen.⁸⁻⁹ Men spreekt van casusspecificiteit: het is niet mogelijk de prestatie van een student in een bepaalde gesprekssituatie (casus) te voorspellen op basis van de prestatie in een andere gesprekssituatie. Dit kan men oplossen met veel verschillende gesprekssituaties, wat praktisch evenwel moeilijk haalbaar is.

Een andere oplossing is om een stations-toets aan te vullen met een videotoets met gespreksfragmenten waarop een reactie wordt gevraagd. Een dergelijke videotoets kan gemakkelijker aan de eisen van betrouwbaarheid en validiteit voldoen.⁹ Een videotoets om communicatievaardigheden te meten wordt in het medisch onderwijs niet vaak aangewend.¹⁰ Men kan ook een schriftelijke toets overwegen. In het medisch onderwijs worden communicatievaardigheden wel vaker gemeten met een schriftelijke toets.¹¹⁻¹³ Meerkeuzevragen zijn hierbij minder geschikt. Een schriftelijke toets met vignetten waarin een gesprekssituatie beschreven wordt, waarop de student dient te reageren alsof hij of zij de gesprekspartner is, bezit een grotere face validity.

In dit onderzoek werd voor drie basiscommunicatievaardigheden (luisteren, vragen stellen en informeren) elk één station voorzien met een gestandaardiseerde gesprekspartner. De drie vaardigheden werden ook gemeten met behulp van schriftelijke en videotoetsen. In de schriftelijke

toets en in de videotoets werd aan de studenten gevraagd een reactie te formuleren alsof ze aan het gesprek deelnamen. Bij de stationstoets waren er instructies voor de student en voor de gesprekspartner, een score-instrument en beoordelingsinstructies; bij de videotoets waren er videofragmenten, een schriftelijke vragenlijst en beoordelingsinstructies en bij de schriftelijke toets een schriftelijke vragenlijst en beoordelingsinstructies. Van de stationstoets werd slechts één versie voorzien. Van de schriftelijke en videotoetsen werden twee combinaties ontwikkeld, die bestonden uit twee inhoudelijk parallelle delen. Van elke vraag werden dus vier varianten geconstrueerd in de hoop na itemanalyse nog drie varianten over te houden, die gebundeld zouden kunnen worden om ze in de toekomst te gebruiken als begin-, tussen- en eindtoets. Door deze werkwijze is de samenstelling van deze toetsen dus gelijkaardig aan de oorspronkelijke, waardoor de inhoudsvaliditeit geen geweld wordt aangedaan.

Aangezien het van meet af aan de bedoeling was een zelfstudiepakket en een communicatietoets samen te stellen die niet enkel geschikt waren voor geneeskundestudenten maar ook voor andere toekomstige hulpverleners, werden beide ook aangeboden aan studenten uit de Faculteit Farmaceutische Wetenschappen. Om de begripsvaliditeit na te gaan werden de toetsen afgenomen bij geneeskundestudenten op twee momenten in de opleiding, zodat de scores van studenten die nog geen communicatietraining gevolgd hadden, konden vergeleken worden met die van studenten die wel een training achter de rug hadden.

Methode

Deelnemers

Aan de stationstoets namen 28 geneeskunde-, 20 huisartsgeneeskunde- en geen

farmaciestudenten deel. Van de schriftelijke en de videotoets bestonden vier versies, waarvan er telkens twee werden aangeboden. De eerste combinatie werd ingevuld door 19 geneeskunde-, 12 farmacie- en 9 huisartsgeneeskundestudenten; voor de tweede combinatie waren de aantallen respectievelijk 15, 13 en 9. De studenten namen vrijwillig deel aan het onderzoek, zodat de resultaten wellicht wat zullen vertekend zijn doordat de meer gemotiveerde studenten hebben meegewerkt. Voor het onderzoek waren 36 studenten geneeskunde, 13 studenten huisartsgeneeskunde en 32 farmaciestudenten benaderd met de vraag om deel te nemen.

Toetsen

De *stationstoets* bestond uit drie stations: luisteren, vragen stellen en informeren. Daarin speelden de gestandaardiseerde gesprekspartners steeds hun rol aan de hand van schriftelijke instructies. Tijdens een voorafgaande introductiesessie hadden zij de schriftelijke instructies ontvangen en de mondelinge richtlijnen over hoe zij moesten reageren op het gedrag van de student. Ook hadden zij de gelegenheid gehad te oefenen waarbij feedback werd gegeven. De student kreeg vóór elk station de betreffende schriftelijke opdracht te lezen. De prestaties van de student werden door twee beoordelaars onafhankelijk van elkaar gescoord aan de hand van beoordelingsinstructies die tevoren met hen waren besproken.

De *videotoets* bestond uit videofragmenten waarbij de acteur zich rechtstreeks tot de kijker richt of tot een in beeld zichtbare gesprekspartner. Voor het acteren werd overwegend een beroep gedaan op leden van een amateurtoneelgezelschap, waarvan sommigen een opleiding dictie hadden gevolgd. Bij de meeste items werd aan de student gevraagd een welbepaalde vaardigheid toe te passen – bijvoorbeeld:

‘Geef in uw eigen woorden de emotionele inhoud van de mededeling weer’ – en te reageren alsof hij of zij als gesprekspartner aan het gesprek deelnam. Om de spontaniteit en natuurlijkheid van de reactie in de hand te werken, werd slechts een beperkte antwoordtijd voorzien, zodat de student onmiddellijk moest reageren.

De *schriftelijke toets* had de vorm van gespreksfragmenten met vragen. Er werd bijvoorbeeld in een paar zinnen een situatie geschetst van een collega die net terug is na een buitenlandse stage van zes maanden, gevolgd door de opdracht: ‘Stel haar twee open vragen in verband met haar stage’. Net als bij de videotoets diende de student bij de meeste items te reageren als gesprekspartner.

De stationsproef bestond uit 17 items, verdeeld over 3 stations. Elke versie van de videotoets bestond uit 10 items en elke versie van de schriftelijke toets uit 12 items. De volgorde van afname was steeds: videotoets, schriftelijke toets, stationstoets. Elk item in iedere toets werd gescoord op een vijfpuntsschaal van ‘slecht’ tot ‘goed’. De stationstoets werd gescoord door huisartsen en gedragswetenschappers. Bij de schriftelijke en videotoetsen werden uitgebreide beoordelingsinstructies gehanteerd met algemene richtlijnen en voorbeelden van goede antwoorden. De beoordelaars hiervan waren docenten verpleegkunde, afgestudeerde apothekers of verpleegkundestudenten die in principe de studenten niet kenden. Bij meerdere beoordelaars werd het rekenkundig gemiddelde als eindscore genomen.

Statistische analyse

Om de psychometrische kwaliteit van de toetsen na te gaan, werden betrouwbaarheid en validiteit berekend. Na afname werden eerst itemanalyses uitgevoerd om

te bepalen of er items verwijderd moesten worden. De betrouwbaarheidsmeting omvatte een berekening van de interne consistentie (Cronbachs alfa) en de interbeoordelaarsbetrouwbaarheid. Voor de interbeoordelaarsbetrouwbaarheid werden de correlaties berekend tussen de scores van alle beoordelaarsduo's die samen ten minste drie studenten hadden beoordeeld. Deze Pearson product-moment-correlatiecoëfficiënten werden gemiddeld na omzetting in Fisher z-waarden.

Voor de validiteit werden de begrips- en de criteriumvaliditeit bestudeerd. Indien studenten die een communicatietraining gevolgd hadden een significant hogere score zouden behalen dan studenten die niet aan een dergelijke training hadden deelgenomen, zou dit een aanwijzing kunnen vormen voor de begripsvaliditeit. De huisartsgeneeskundestudenten hadden deze training gevolgd, de geneeskundestudenten niet. Verschillen tussen beide groepen werden op hun statistische significantie getoetst met behulp van eenzijdige t-toetsen. De criteriumvaliditeit zou blijken uit het ontbreken van significante correlaties tussen de verschillende toetsvormen. Hiervoor werden Pearson product-moment-correlatiecoëfficiënten berekend.

Resultaten

Itemanalyse

Voor eliminatie kwamen vooral items met een gemiddeld hoge of lage score in aanmerking. Ook als het schrappen van een item leidde tot een hogere betrouwbaarheid van de toets – gemeten met Cronbachs alfa – werd eliminatie overwogen. Deze itemanalyse mondde uit in drie versies van zowel de schriftelijke als de videotoets. Elke versie van de videotoets bevatte 9 items en elke versie van de schriftelijke toets 12 items.

Betrouwbaarheid

Cronbachs alfa bedroeg 0.88 bij de stationsproef en varieerde van 0.46 tot 0.67 voor de videotoets en van 0.58 tot 0.69 voor de schriftelijke toets. De interbeoordelaarsbetrouwbaarheid was behoorlijk (ongeveer gelijk aan of groter dan 0.70) voor de stations over informeren en luisteren en voor videotoets 1 en 2 en schriftelijke toets 1. Voor het station vragen stellen en voor videotoets 3 en de schriftelijke toetsen 2 en 3 was de interbeoordelaarsbetrouwbaarheid lager (tabel 1).

Validiteit

Voor de stationstoets, de videotoetsen 1 en 3 en de schriftelijke toetsen 2 en 3 hadden de huisartsgeneeskundestudenten een significant hogere gemiddelde score dan de geneeskundestudenten (tabel 2). Wat de criteriumvaliditeit betreft werden alleen significante verbanden gevonden tussen videotoets 1 en 2 en tussen de schriftelijke toetsen 2 en 3. De correlaties tussen de videotoetsen, schriftelijke toetsen en de stationstoets waren niet significant (tabel 3).

Bespreking

Louter op theoretische gronden is een combinatie van toetsen voor basiscommunicatievaardigheden, zoals ontwikkeld in het kader van dit onderzoek, veelbelovend. Niet alleen bestaat de stationsproef uit drie stations, terwijl in gelijkaardige onderzoeken vaak slechts één station wordt gebruikt, maar bovendien kan deze stationstoets nog worden aangevuld met een schriftelijke en een videotoets. Een eindoordeel gebaseerd op meer toetsvormen is betrouwbaarder dan een oordeel op basis van één toetsvorm.¹⁴ Verder verschillen deze toetsen van de gangbare doordat de communicatievaardigheden los van een medische inhoud worden gemeten. Dit maakt de toets ook poten-

Tabel 1. *Interne consistentie en interbeoordelaarsbetrouwbaarheid van de toetsen afgenomen bij studenten geneeskunde (G), farmacie (F) en huisartsgeneeskunde (H).*

	Aantal studenten totaal	(G+F+H)	Aantal items	Cronbachs alfa	Correlatie tussen beoordelaars
Stationstoets	48	(28+0+20)	17	0.88	0.69
- Informeren			6		0.79
- Luisteren			7		0.77
- Vragen stellen			4		0.40
Videotoets 1	40	(19+12+9)	9	0.67	0.77
Videotoets 2	40	(19+12+9)	9	0.46	0.68
Videotoets 3	37	(15+13+9)	9	0.54	0.65
Schriftelijke toets 1	40	(19+12+9)	12	0.58	0.84
Schriftelijke toets 2	37	(15+13+9)	12	0.67	0.39
Schriftelijke toets 3	37	(15+13+9)	12	0.69	0.43

Tabel 2. *Vergelijking gemiddelde toetsscores (gem.) met standaarddeviaties (sd.) van geneeskunde- en huisartsgeneeskundestudenten.*

	Maximale score	Studenten geneeskunde		Studenten huisartsgeneeskunde		Statistische significantie
		gem.	(sd.)	gem.	(sd.)	
Stationstoets	85	53.3	(10.0)	64.0	(5.9)	p < 0.001
Videotoets 1	45	20.9	(4.4)	26.6	(7.8)	p < 0.05
Videotoets 2	45	21.7	(4.7)	23.2	(4.8)	n.s.
Videotoets 3	45	23.5	(3.8)	29.6	(3.0)	p < 0.001
Schriftelijke toets 1	60	39.7	(7.0)	42.1	(6.6)	n.s.
Schriftelijke toets 2	60	42.0	(5.9)	48.4	(3.6)	p < 0.05
Schriftelijke toets 3	60	40.4	(6.2)	49.1	(2.9)	p < 0.01

tieel nuttig voor andere onderwijsinstellingen waar hulpverleners worden opgeleid. Ten slotte pogen deze toetsen communicatiegedrag en niet kennis omtrent communicatie te meten. Niet alleen in de stationstoets maar ook in de schriftelijke en de videotoets werd de studenten gevraagd te reageren alsof zij deelnemer

waren aan het gesprek waarvan zij een fragment te zien of te lezen kregen.

Het bleek mogelijk na itemanalyse drie versies over te houden van de oorspronkelijke vier versies van de schriftelijke en de videotoetsen. Tegen de verwachting in was de stationstoets het meest betrouwbaar, zowel naar interne consistentie als

Tabel 3. *Correlaties tussen de scores op de videotoetsen (VT1 t/m 3), de schriftelijke toetsen (ST1 t/m 3) en de stationstoets (RT).*

	VT1	VT2	ST1	RT	VT3	ST2	ST3
VT1							
VT2		0.51*					
ST1	0.28	0.13					
RT	0.27	0.07	0.21		0.24	-0.17	0.06
VT3						0.17	0.14
ST2							0.75*
ST3							

* $p < 0.001$

naar interbeoordelaarsbetrouwbaarheid. De interbeoordelaarsbetrouwbaarheid van sommige toetsen (VT1, VT2, ST1) was behoorlijk tot goed. De lage interbeoordelaarsbetrouwbaarheid bij de schriftelijke toetsen 2 en 3 zou kunnen te wijten zijn aan de toevallige samenstelling van de proef- en/of beoordelaarsgroep. Voor de praktijk betekent dit dat de stationstoets betrouwbaarder is dan de schriftelijke en de videotoetsen.

De bevinding dat de huisartsgeneeskundestudenten hogere scores behaalden dan de geneeskundestudenten bij vijf van de zeven toetsen is een aanwijzing voor begripsvaliditeit. Een kanttekening is dat de huisartsstudenten drie jaar verder zijn met hun studie dan de geneeskundestudenten. Het lijkt mogelijk met de toetsen prestatieverschillen te detecteren in groepen die van elkaar verschillen in het al dan niet gevolgd hebben van meer onderwijs waaronder een communicatietraining. De correlaties tussen de stationstoets, de schriftelijke toetsen en de videotoetsen onderling waren niet significant. Dit wekt de indruk dat ze verschillende aspecten van communicatievaardigheid meten. Een betrouwbaar en valide

meetinstrument voor basiscommunicatievaardigheden moet daarom bestaan uit verschillende toetsen.

Aangezien de toetsen doelbewust werden ontwikkeld los van een medische inhoud, zijn ze niet geschikt om *medische* communicatievaardigheden te meten. Dit vergt de samenstelling van gelijkaardige toetsen die – met het oog op de casusspecificiteit – betrekking hebben op zoveel mogelijk verschillende medische probleemsituaties.

Dankwoord

Wij danken Jo Goedhuys, Manu Keirse, Lies Leemans, Rob Stroobants en Chantal van Audenhove voor hun advies bij de ontwikkeling van de toetsen.

Dit onderzoek werd uitgevoerd in het kader van het onderwijsgerichte onderzoeks-, ontwikkelings- en implementatieproject (OOI) 'Effectmeting van een begeleide zelfinstructie voor elementaire sociale vaardigheden bij studenten geneeskunde en apotheker', dat wordt gesubsidieerd door de Onderwijsraad van de KU Leuven.

Literatuur

1. Beullens J, Rethans JJ, Goedhuys J, Buntinx F. The use of standardized patients in research in general practice. *Family Practice* 1997;14:58-62.
2. Brown JE, O'Shea JS. Improving medical student interviewing skills. *Pediatrics* 1980;65:575-8.
3. Evans BJ, Stanley RO, Mestrovic R, Rose L. Effects of communication skills training on students' diagnostic efficiency. *Med Educ* 1991;25:517-26.
4. Mason JL, Barkley SE, Kappelman MM, Carter DE, Beachy WV. Evaluation of a self-instructional method for improving doctor-patient communication. *J Med Educ* 1988;63:629-35.
5. Vaughan M, Marks JN. Teaching interviewing skills to medical students: a comparison of two methods. *Med Educ* 1976;10:170-5.
6. Weihs K, Chapados JT. Interviewing skills training: a study. *Soc Sci Med* 1986;23:31-4.
7. Usherwood T. Subjective and behavioural evaluation of the teaching of patient interview skills. *Med Educ* 1993;27:41-7.
8. Hodges B, Turnbull J, Cohen R, Bienenstock A, Norman G. Evaluating communication skills in

- the objective structured clinical examination format: reliability and generalizability. *Med Educ* 1996;30:38-43.
9. Smit GN. De beoordeling van professionele gespreksvaardigheden: Constructie en evaluatie van rollenspel-, video- en schriftelijke toetsen. [Proefschrift]. Groningen: Rijksuniversiteit Groningen; 1995.
 10. Kendrick T, Freeling P. A communication skills course for preclinical students: evaluation of general practice teaching using group methods. *Med Educ* 1993;27:211-7.
 11. Garrud P, Chapman IR, Gordon SA, Herbert M. Non-verbal communication: evaluation of a computer-assisted learning package. *Med Educ* 1993;27:474-8.
 12. Knox JDE, Bouchier IAD. Communication skills teaching, learning and assessment. *Med Educ* 1985;19:285-9.
 13. Winefield HR. Subjective and objective outcomes of communication skills training in first year. *Med Educ* 1982;16:192-6.
 14. Starren J, Bakker SJ, van der Wissel A. Inleiding in de onderwijspsychologie. Muiderberg: Dick Coutinho; 1988.

De auteurs:

J. Beullens is psycholoog en wetenschappelijk medewerker aan de Onderwijskundige Dienst van de Faculteit Geneeskunde van de Katholieke Universiteit Leuven.

C. Aubry is licentiaat verplegingswetenschappen en coördinator van het Vaardigheidscentrum van de Faculteit Geneeskunde van de Katholieke Universiteit Leuven.

Prof. Dr. H. Jaspaert, pedagoog, is hoofddocent en leidt de Onderwijskundige Dienst van de Faculteit Geneeskunde van de Katholieke Universiteit Leuven.

Correspondentieadres:

Johan Beullens, Onderwijskundige Dienst Faculteit Geneeskunde, KU Leuven, Minderbroedersstraat 17, B-3000 Leuven, België, e-mail: Johan.Beullens@med.ku-leuven.ac.be.

Summary

Introduction: Three tests were designed for the assessment of basic communication skills to evaluate a self-study pack for different training programmes in (health) care professions. The tests had no specific medical content.

Method: For each basic dimension (listening, asking questions and informing) a station of an objective structured clinical examination (OSCE) was developed where the student had an encounter with a standardised discussion partner. In addition there were four versions of a video test and a written test. The tests were administered to medical students (fourth year), pharmacology students (fourth year) and general practice students (seventh year). Item analysis resulted in three versions of the video and the written tests.

Results: The OSCE appeared to be more reliable than the three versions of the video and written tests, respectively. General practice students, who had attended communication skills training, scored significantly higher on five out of the seven tests compared with the fourth year students who had not attended such training. Correlation between the different types of tests was limited.

Conclusions: The higher scores of the GP students are an indication of test validity. The limited correlation between the different types of tests indicates that the tests measure different aspects of communication skills. A reliable and valid assessment of communication skills requires different types of tests. The tests used measure basic communication skills. In order to test specific medical communication skills, tests should be designed with medical content. (Beullens J, Aubry C, Jaspaert H. Psychometric qualities of three tests for basic communication skills. *Dutch Journal of Medical Education* 2001;20(2):54-60.)